

Project Title: Breaking the Internet Barrier

IDRC project number: 104519-006 and 014

Research organizations involved in the study:

Tsinghua University, China, and University of
Waterloo, Canada

Location of Study: Beijing China. Waterloo, Ont.

By: Xiaoyan Zhu (Tsinghua University)

Ming Li (University of Waterloo)

Date: March 5, 2017

- © 2017 Ming Li and Xiaoyan Zhu
- Disseminated under creative commons attribution
license (<http://creativecommons.org/licenses/by/4.0/>)

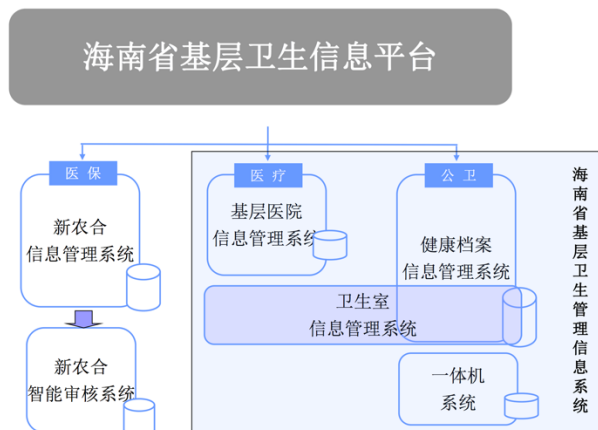
Since in the first project we have established solid theories (for instance, conditional information distance, semantic distance, concept relatedness, etc.) and techniques (such as question analysis, answer extraction, answer summarization) on question answering, the major efforts in this project are paid to developing real-world applications empowered by our key algorithms and techniques. We have developed or participated in developing three applications:

- A medical information management system which integrates clinical, healthcare, medical insurance, and hospital information at rural areas. The system has been distributed to clinical and health departments of Hainan Province, and is serving more than 6 million users.
- A clinical information system that assists physicians to make better informed decision at Point of Care, and can deliver relevant and evidence-based medical information to their patients.
- A chatting robot named Doudou that is known for its high ability of chatting with kids. Doudou is already sold on JD.com (equivalent to Amazon in China), and will be used in Chinese kindergardens.

1. Medical Information Management System

Our question answering techniques are successfully applied to a medical information

management system which is distributed over the clinical and health departments of Hainan Province. As shown in the figure, the system is a medical information management system which integrates clinical, healthcare, medical insurance, and hospital information at rural areas. Our question answering techniques empower the system with a natural language interface to query such various information from databases, documents, and so on.



Overview of the Medical Information Management System in Hainan Province.

The features of the system lies in several folds:

- 1) It's online available and easy to access through web service and Wechat service.
- 2) It can be used by either professional doctors or ordinary patients. Moreover, it can be used by less-trained doctors for more accurate prescription.
- 3) It offers accurate question answering results which is empowered by our question answering techniques.

Since released in 2015, the system has been installed to more than 50 medical institutes at about 20 districts, and serves more than 6.54 million outpatients, 23900 inpatients, and 23700 others.

In addition, on top of the system, Hainan Province has established expert websites and Wechat interface to share and access clinical and medical information.

The system has been operating for more than two years, and serving more than 6 million users.

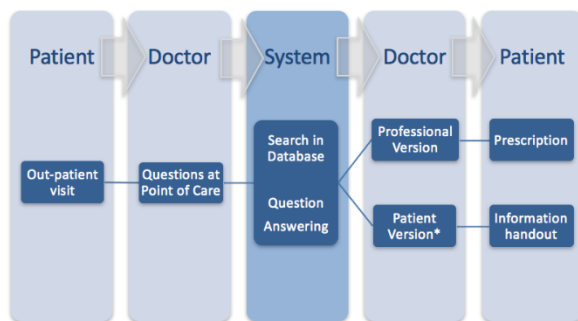
2. Look4e: A Supporting System for Chinese Doctors

We helped a company named Look4e to develop a clinical information system to assist physicians to make better informed decision at Point of Care. And through physicians, the system can deliver relevant and easy to understand evidence-based medical information to their patients. The system is mainly designed to solve the information overload problems that physicians are faced up with:

- 1) Clinical knowledge is doubled every 18 months; Physicians need to read 19 professional journal articles daily to keep up-updated; But physicians are just too busy to keep up with.
- 2) Large number of out-patient limits the time to spend with patient;

misleading/biased information creates more difficulties in communication with physicians.

The system works as follows, as shown in the below figure: the system provides a professional version which will be used by doctors for prescription, and a patient version which will be used by patients for their access to clinical information. In both versions, the system is equipped with two functionalities: one is direct search in medical/clinical databases, while the other is a natural interface through which doctors or patients can type natural language questions and fetch back a direct answer. The natural language interface is powered by our techniques. This interface requires our question answering engine to parse clinical questions and find an accurate answer in either structured clinical database or unstructured medical publications.



The work flow of Look4e.

Look4e is commercialized and successfully applied to Chinese doctors and patients. By July of 2017, the system will be used in about 100 hospitals in China, and at the end of this year, it will be used in more than 500 hospitals.

3. Chatting Robot: Doudou

Other than the medical information system, we have also commercialized the chatting Robot Doudou. Doudou speaks Chinese, is already sold on JD.com (equivalent to Amazon in China). Doudou is known for its high ability of chatting with kids. This chatting robot serves several roles, as follows:

- 1) It is a prototype for a future AI technology of being able to speak. Doudou is currently the platform that has the best ability of chatting with people. Many robots in China are using Doudou platform. Its goal is eventually serving in intelligent homes, smart cars, as well as automatic answering services.
- 2) Doudou is equipped currently with a lot of children educational materials. In China's rural areas, many schools have just one teacher, who teaches everything from Chinese to math. Doudou educational assistant aims to become a teaching assistant for these children.
- 3) Doudou can accompany the elderly.
- 4) Doudou's next project is learning English, provide English learning kids a "native English environment". Such an approach is only made possible by chatting robots.

The technology was licensed to RSVP Technologies Inc. This is our startup company. The technologies are not owned by RSVP Technologies. The company has attracted investment of a few million Canadian dollars. It has produced the Doudou robot in the Chinese market, partially using our technology (QA part). The Doudou robot is created for Chinese children, and in the future, it will help them to learn English. Doudou writes poems, talks to kids, and is now learning to write composition. Although the sales data is confidential information and hence cannot be shared, the Doudou robot has attracted attention from major companies like Huawei, XiaoMi, Haier, and HuiXiang (and China Mobile company that plans to use Doudou for 50,000 Chinese kindergardens). The attached photos show Doudou on Wall Street, and on Chinese CCTV (with Dr. Li).

Trainees of the last 2 years

List of Trainees

	Trainees supervised by IDRC RC	Trainees supervised by the CRC	Trainees co-supervised by IDRC RC & CRC	Trainees supervised by other collaborators
Undergraduate	Juan Du, Mingdong Wang, Wentao Dai, Hao Zhou, Ye Zhang, Tianjian Jiang, Mingzhao Zhang, Han Xu			
Masters	Yipeng JIANG, Yicheng LIU, Biao LIU, Xiaoji DING, Qingyang XIA	Yahui Chen (from Tsinghua), Borui Ye (from Tsinghua), Kun Xiong (from Tsinghua)		
Doctoral	LI ZHAO, Jun FENG, Han XIAO, Lei FANG, Shouzhong TU, Daoyi LI, Yequan WANG, Zheng ZHANG	Guangyu Feng (from Tsinghua)		
Post-doctoral	Bin LIU, Cheng LING, Ting Wang	Xiaopeng Yang, Anqi Cui (from Tsinghua), Nancy		

		Zhang, Hieu Ngoc Tran.		
Other				

Cumulative list of research outputs in the during the 2nd period

Type	Total number of Research Outputs
Journal Articles (published / accepted)	6
Conference Papers	17
Presentations (non-academic)	8
Books	
Book Chapters	
Newspapers / other media	
Theses	10
- MA / MSc	5
- PhD	5

All research outputs are included in the bibliography below.

Bibliography

Research Output Bibliography

*Citation format: author(s), date, title, publisher, active weblink if available)

Journal Articles (published/accepted)

1. Chong Long, Jie Zhang, Minlie Huang, Xiaoyan Zhu, Ming Li, Bin Ma. 2014. Estimating feature ratings through an effective review selection approach. Knowl. Inf. Syst. 38(2): 419-446 (2014)
2. Po Hu, Minlie Huang, Xiaoyan Zhu. 2014. Journal of Computer Science and Technology, Vol.29, Num.3, May 2014
3. Po Hu, Minlie Huang, Xiaoyan Zhu. 2014. Patent Key Component Extraction with the Application of Patent Similarity Analysis. Journal of Computational Information Systems (2014) 5813 – 5820
4. Minlie Huang, Qiao Qian, Xiaoyan Zhu. Encoding Syntactic Knowledge in Neural Networks for Sentiment Classification. Accepted by ACM transaction on information system.

Conference Papers

1. Yequan Wang, Minlie Huang and xiaoyan Zhu. *Attention-based LSTM for Aspect-level Sentiment Classification*. EMNLP 2016, Austin, Texas, USA.

2. Jun Feng, Minlie Huang, Yang Yang, Xiaoyan Zhu. GAKE: Graph Aware Knowledge Embedding. COLING 2016, Osaka, Japan.
3. Hao Zhou, Minlie Huang, Xiaoyan Zhu. Context-aware Natural Language Generation for Spoken Dialogue Systems. COLING 2016, Osaka, Japan.
4. Naitong Yu, Minlie Huang, Yuanyuan Shi, Xiaoyan Zhu. Product Review Summarization by Exploiting Phrase Properties. COLING 2016, Osaka, Japan.
5. Biao Liu, Minlie Huang, et al. A Sentence Interaction Network for Modeling Dependence between Sentences. ACL 2016, Berlin, Germany.
6. Han Xiao, Minlie Huang, Xiaoyan Zhu. TransG : A Generative Model for Knowledge Graph Embedding.. ACL 2016, Berlin, Germany.
7. Han Xiao, Minlie Huang, Xiaoyan Zhu. From One Point to A Manifold: Knowledge Graph Embedding For Precise Link Prediction. IJCAI 2016, New York, USA.
8. Li Zhao, Minlie Huang, et al. Semi-Supervised Multinomial Naive Bayes for Text Classification by Leveraging Word-Level Statistical Constraint. AAAI 2016, Phoenix, Arizona, USA.
9. Qiao Qian, Bo Tian, Minlie Huang et al. Learning Tag Embeddings and Tag-specific Composition Functions in Recursive Neural Network. ACL 2015, Beijing, China.
10. Biao Liu, Minlie Huang, et al. Incorporating Domain and Sentiment Supervision in Representation Learning for Domain Adaptation. IJCAI 2015, July 25-31 Buenos Aires, Argentina.
11. Li zhao, Minlie Huang, Xiaoyan Zhu. Sentiment Extraction by Leveraging Aspect-Opinion Association Structure. CIKM 2015, Oct 19-23, Melbourne, Australia.
12. Li Zhao, Minlie Huang, Haiqiang Chen, Junjun Cheng, Xiaoyan Zhu. Clustering Aspect-related Phrases by Leveraging Sentiment Distribution Consistency. EMNLP 2014, October 25-29, 2014 – Doha, Qatar.
13. Lei Fang, Qiao Qian, Minlie Huang, Xiaoyan Zhu. Ranking Sentiment Explanations for Review Summarization Using Dual Decomposition. CIKM 2014. Nov 3-7, 2014, Shanghai, China.
14. Chong Long, Jie Zhang, Minlie Huang, Xiaoyan Zhu, Ming Li, Bin Ma: Estimating feature ratings through an effective review selection approach. Knowl. Inf. Syst. 38(2): 419-446 (2014)
15. Lijing Qin, Shouyuan Chen, Xiaoyan Zhu. 2014. Contextual Combinatorial Bandit and its Application on Diversified Online Recommendation. In Proceedings of SDM 2014, ; Best Student Paper runner up, Philadelphia, Pennsylvania, USA
16. Minlie Huang, Borui Ye, Yichen Wang, Xiaoyan Zhu. New Word Detection for Sentiment Analysis. ACL 2014, June 23-25, 2014. Baltimore, Maryland, USA.

Presentation (non-academic)

1. Tianyi Public Lecture, Ningbo City, 300 people. Ming Li: “Latent Search”
2. Keynote Speech: Chatting robots by deep learning. Global Artificial Intelligence and Robotics Summit, Chinese Computer Federation, Shenzhen, China, August 12-13, 2016.
3. Keynote Speech: Chatting robots by deep learning. 7th Workshop on data mining and intelligent computing. Hefei, China, August 9-10, 2016.

4. Keynote Speech: Chatting robots by deep learning. The 3rd workshop on big data and computational intelligence. Beijing, July 29-July 31, 2016.

Ph.D

1. Lijing Qin, Statistical Models and Learning Algorithms for Recommender System
2. Xingwei Zhu, The Structured Organization of Social Media Contents and its Applications
3. Li Zhao, Sentiment Analysis with Data-driven Constraints
4. Dan Holtby, LOOPWEAVER: Protein Loop modeling
5. Guangyu Feng, Approximating semantics, Defense 2017 March 10, expected.

Master

1. Tanche Li, A Chinese Question Answering System for Specific Domain
2. Yicheng Liu, A Question Answering System for Specific Domain
3. Yipeng Jiang, A Question Answering System for Customer Service of Mobile Products
4. Kun Xiong, Similarity metric for natural language processing

Media coverage.

The advertisement of Doudou posted on Wall Street



Dr. Ming Li was interviewed by CCTV (China's central TV) for Doudou robot, on January 27, 2017.

